**Final Report for Period:** 07/2011 - 12/2011
**Submitted on:** 12/02/2011
**Principal Investigator:** Wu, C. F. Jeff .
**Award ID:** 0705261
**Organization:** Georgia Tech Research Corp
**Submitted By:**
Wu, C. F. Jeff - Principal Investigator
**Title:**
Collaborative Research: GOALI Statistical Methods for Modern IT Systems

## Project Participants

**Senior Personnel**

**Name:** Wu, C. F. Jeff
**Worked for more than 160 Hours:** Yes
**Contribution to Project:**

**Post-doc**

**Graduate Student**

**Name:** Hung, Ying
**Worked for more than 160 Hours:** No
**Contribution to Project:**

**Name:** Deng, Xinwei
**Worked for more than 160 Hours:** Yes
**Contribution to Project:**

**Name:** Adiga, Nagesh
**Worked for more than 160 Hours:** No
**Contribution to Project:**

**Name:** Tan, Matthias
**Worked for more than 160 Hours:** Yes
**Contribution to Project:**

**Name:** Wang, Yijie
**Worked for more than 160 Hours:** Yes
**Contribution to Project:**

**Undergraduate Student**

**Technician, Programmer**

**Other Participant**

**Research Experience for Undergraduates**

## Organizational Partners

## Other Collaborators or Contacts

Professor Peter Qian is the collaborator who holds a separate grant at U of Wisconsin. Dr. Y. Amemiya is a coPI from IBM who does not receive support but is actively involved in the collaborative work. Other collaborators are Professors M Ai (Peking U), A. Mandal (U of Georgia), Y. Hung (Rutgers), T Dasgupta (Harvard U), W Wang (Nat Taiwan U), R B Chen (Nat Kaohsiung U), and R. Joseph (Georgia Tech).

## Activities and Findings

**Research and Education Activities:**

The project has made very good progress. In four years, the PI has produced 11 papers: ten have appeared; all in major journals; and one to appear as a book chapter. Five students were recruited to work on the project, some under partical support. The work has been presented in national meetings. The PI has met with the other PIs (Qian and Amemiya) on several occassions (visit to IBM, hosting their visit to Georgia Tech, conferences in Madison and in Atlanta) to discuss and exchange our research results.

**Findings:**

In paper 1 a new hierarchical Bayesian approach is proposed that combines data from two sources of computer experiments: one with high accuracy and cost, the other with low accuracy and high cost. The latter has a larger sample while the former has a smaller sample but higher accuracy. It employs some novel techniques in the Bayesian computations. Paper 7 addresses the accompany design question of constructing Latin hypercubes that can accommodate experiments with two levels of accuracy. Paper 2 solves an outstanding problem of finding Gaussian process models for computer experiments with both quantitative and qualitative factors. Such experiments are abundant in practice and the existing kriging method cannot be applied. Paper 6 addresses the accompany design question of constructing Latin hypercubes that can accommodate experiments with both quantitative and qualitative factors. Papers 6 and 7 make some fundamental contributions to the theory of designs. Paper 3 gives a general approach to finding optimal blocking plans that incorporate the foldover techniques for two-level designs. Paper 4 continues the novel Bayesian approach developed by Joseph to find optimal mixed two-and four-level designs. Its motivation and derivations are Bayesian but the final criterion is frequentist and resembles the minimum aberration criterion. Paper 5 expands on the SELC technique of Mandal and Wu by bringing the kriging method to help accelerate the SELC's search for optima. It outperforms both SELC and a kriging-based optimization method. Paper 8 proposes a new method to build surrogate models for computer experiments by using over-complete basis in applied math. It should find general applications in industries like IBM. Paper 9 gives an in-depth analysis of Shanin's variable search method, which forms the backbone of the well-known Shanin's method for quality improvement. Shanin's method has been widely used in industries but lacks any theoretical justification. In paper 10 a new class of designs called probability-based Latin hypercube designs is proposed and methods of contruction given. It is based on work done at IBM (second author is from IBM) to address a practical question on how to allocate sensors in a data center. In paper 1 under book chapter, the two PIs and Dr. Hung wrote a review article on statistical design and analysis for data center thermal management. It will appear in a special volume on data center problems.

**Training and Development:**

Ms. Hung, who received partial support in the summer, has been teaching as a tenuer track assistant professor at the Rutgers Statistics Department. Mr. Deng graduated in summer of 2009, taught at the Stat Dept, U of Wisconsin Madison for two years, and has joned Virginia Tech stat dept as a tenure track assistant professor. Since this project is academic-industrial in nature, it offers Wu a unique opportunity to tackle practically important and statistically challenging problems from the IT industry. The methods developed are also found to be useful for solving pressing issues faced by the IT industry like slashing energy consumption of IT systems. The materials developed in this project will be used in teaching his popular classes on design and analysis of experiments at Georgia Tech.

**Outreach Activities:**

The PI has been giving seminars and talks at universities and conferences to different audiences on topics related to this project. All of these activities improved public understanding and awareness of science and technology, especially IT.

## Journal Publications

P. Z. G. Qian and C. F. J. Wu, "Bayesian hierarchical modeling for integrating low-accuracy and high-accuracy experiments", Technometrics, p. , vol. 50, (2008). Published,

P. Z. G. Qian, H. Wu, and C. F. J. Wu, "Gaussian process models for computer experiments with qualitative and quantitative factors", Technometrics, p. , vol. 50, (2008). Published,

M. Ai, X. Xu and C. F. J. Wu, "Optimal blocking and foldover plans for regular two-level designs", Statistica Sinica, p. , vol. 20, (2010). Published,

V. R. Joseph, M. Ai and C. F. J. Wu, "Bayesian-inspired optimal mixed two- and four-level designs", Biometrika, p. , vol. 96, (2009). Published,

A. Mandal, P. Ranjan and C. F. J. Wu, "G-SELC: optimization by sequential elimination of level combinations using genetic algorithms and gaussian processes", Ann. Applied Statistics, p. , vol. 3, (2009). Published,

P. Z. G. Qian and C. F. J. Wu, "Sliced space-filling designs", Biometrika, p. 9, vol. 96, (2009). Published,

P. Z. G. Qian, M. Ai and C. F. J. Wu, "Construction of nested space-filling designs", Annals of Statistics, p. , vol. 37, (2009). Published,

R. B. Chen, W. Wang and C. F. J. Wu, "Building surrogates with overcomplete bases in a computer experiment for bistable laser diodes.", IIE Transactions on Quality and Reliability, p. , vol. 43, (2011). Published,

T. Dasgupta, N. Adiga and C. F. J. Wu, "A closer look at Dorian Shainin's variable search technique.", J Quality Technology, p. , vol. 43, (2011). Published,

Y. Hung, Y. Amemiya, and C. F. J. Wu, "Probability-based Latin hypercube designs", Biometrika, p. 96, vol. 97, (2010). Published,

## Books or Other One-time Publications

Y. Hung, P. Qian, and C. F. J. Wu, "Statistical design and analysis in data center thermal management", (2011). chapter in book, Accepted
Editor(s): Y. Joshi and P. Kumar
Collection: Energy Efficient Thermal Management of Data Centers
Bibliography: Springer

## Web/Internet Site

## Other Specific Products

## Contributions

**Contributions within Discipline:**

The statistical methods developed have some major impacts on the design and analysis of computer models and on statistical modeling for IT systems. Two previously unsettled important problems have been solved: the development of the first statistical framework for building Gaussian process models with qualitative and quantitative factors (paper 2) and the introduction of the first systematic work on design construction for computer experiments with different levels of accuracy (papers 1 and 7). The proposed method 'Probability-based Latin hypercube designs' (paper 10) has been adopted at IBM.

**Contributions to Other Disciplines:**

**Contributions to Human Resource Development:**

**Contributions to Resources for Research and Education:**

Our success story of developing novel statistical methods for solving modern IT problems can be used as teaching materials for courses and training classes related to design of experiments and stochastic simulations.

**Contributions Beyond Science and Engineering:**

Some of the developed statistical methods have been shown to be useful for improving thermal management and cooling efficiency of large computer systems, thus offering great potential for reducing IT energy consumption.

**Conference Proceedings**

**Categories for which nothing is reported:**

Organizational Partners

Any Web/Internet Site

Any Product

Contributions: To Any Other Disciplines

Contributions: To Any Human Resource Development

Any Conference